PILOT STUDIES OF THE GENETICS OF OBESITY IN THE WESTERN ALASKA NATIVE POPULATION

By

Erik Briggs Harrington

RECOMMENDED:

Advisory Committee Chair

Head, Program in Biochemistry and Molecular Biology

APPROVED:

Dean, College of Science, Engineering, and Mathematics

Dean of Graduate School

April 30, 2004

Date

PILOT STUDIES OF THE GENETICS OF OBESITY IN THE WESTERN ALASKA NATIVE POPULATION

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Erik Briggs Harrington, B.S.

Fairbanks, Alaska

May 2004

ABSTRACT

Obesity in Alaska Natives is increasing, posing significant health risk for the development of associated diseases. This study examined candidate obesity genes in a set of anonymized Alaska Native DNA samples for loci that might predict obesity risk. DNA samples were divided into three groups according to body mass index: lean (BMI $\leq$ 23), in-between (23< BMI <30), and obese (BMI $\geq$ 30). Screening of 5043 base pairs from the exons of leptin (LEP), leptin receptor (LEPR), neuropeptide-y (NPY), and melanocortin-4-receptor (MC4R) yielded previously discovered SNPs in NPY and LEPR. Additionally, two known promoter region SNPs in NPY and Uncoupling Protein-2 (UCP2) were analyzed. SNPs were in Hardy Weinberg equilibrium, showed little genetic variation between populations, and were not associated with BMI category. We concluded that the study lacked power to detect an association due to an inability to correct for variables known to affect BMI and a small sample size. However, this study enabled pilot studies of several SNP genotyping platforms and the evaluation of allele frequencies in the Alaska Native population, illuminating the value of known SNP genotyping vs. SNP discovery and the benefit of a direct association study vs. an indirect association study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank Dr. Bert Boyer for making this study possible, providing guidance, and allowing a significant amount of research independence. I greatly appreciate Dr. Gabriela Antunez de Mayolo Wilmking and Laura Brunner's work genotyping the UCP2 and NPY promoter SNPs. The National Institutes of Health funded this research from the National Center for Research Resources.

GENERAL INTRODUCTION

OBESITY CONCERN IN THE ALASKA NATIVE POPULATION

Obesity and diabetes have become significant health concerns in the Alaska Native population with the infusion of a westernized "obesogenic" environment (French *et al.*, 2001; Hill and Peters, 1998; Mouratoff *et al.*, 1967; Mouratoff and Scott, 1973). In pre-contact Alaska there was virtually no obesity in the Native population (Michael, 1967). A study by Mouratoff *et al.* showed increasing rates of obesity in the Alaska Native population between 1962 and 1972 (Mouratoff *et al.*, 1967; Mouratoff and Scott, 1973). Non-insulin-dependant diabetes mellitus was shown to be rare in 1957 and is now increasing (Mouratoff *et al.*, 1967; Scott and Griffith, 1957). There is evidence that these changes may be fueled by increased consumption of calorie rich non-indigenous food and a decrease in physical activity associated with mechanization (Adler *et al.*, 1996; Murphy *et al.*, 1995). One of the most notable dietary changes is the increase in carbohydrate intake from 3% to 5% in pre-westernized times to 50% of total energy in 1978 (Heller and Scott, 1967; Knapp and Panruk, 1978; Murphy *et al.*, 1995).

AIM OF STUDY

The aim of this study is to discover single nucleotide polymorphisms (SNPs) in candidate obesity genes that are informative for obesity risk in the Alaska Native population. Body weight is a complex condition largely affected by the interaction of environment and genetics (Barsh *et al.*, 2000; Loos and Bouchard, 2003). The portion of trait variation described by genetic transmission, heritability, for obesity has been shown to be around 70% in twin studies and 50% in familial studies (Allison *et al.*, 1996; Luke *et al.*, 2001; Maes *et al.*, 1997; Stunkard *et al.*, 1990). Familial studies have shown that there is double the risk of developing obesity (BMI $\geq$ 30) when a genetically related family member is obese when compared to the risk of obesity in the whole population; the relative risk is greater when a family has extreme obesity (BMI $\geq$ 45) (Lee *et al.*, 1997). Genome scans of families have identified chromosomal regions that co-segregate with obesity but no major obesity genes affecting a large proportion of society have been identified (Adeyemo *et al.*, 2003; Arya *et al.*, 2004; Comuzzie *et al.*, 1997; Loos and Bouchard, 2003).

PHYSIOLOGY OF CANDIDATE GENE PRODUCTS

This study focuses on 5 candidate genes for obesity: leptin (LEP), leptin receptor (LEPR), neuropeptide y (NPY), melanocortin 4 receptor (MC4R), and uncoupling protein 2 (UCP2). These genes have been implicated as candidate obesity genes by their genetic associations with body weight and by supporting knowledge of the physiology of their gene products. The arcuate nucleus (AN) and paraventricular nucleus in the hypothalamus have been identified as the areas of the brain most responsible for energy regulation. The hypothalamus is where the gene products of LEP, LEPR, MC4R, and NPY act in pathways regulating energy.

Leptin is an endocrine hormone secreted by adipocytes. It signals long-term energy supply by binding to leptin receptors in the hypothalamus, stimulating catabolism and inhibiting anabolism. Leptin binds to the leptin receptor activating janus kinases (JAK) that phosphorylate signal transduction and transcription (STAT) proteins (Schwartz *et al.*, 2000). STAT proteins facilitate the transcription of specific gene targets, presumably related to energy homeostasis.

The **leptin receptor** is a cytokine receptor whose activation stimulates expression of suppressor of cytokine signalling-3 (SOCS-3) proteins that inhibit further activation of the by leptin (Bjorbaek *et al.*, 1998). Leptin deficiency causes severe obesity in ob/ob mice (Zhang *et al.*, 1994).

Neuropeptide Y is a widely expressed neurotransmitter that affects energy homeostasis through its action in the arcuate nucleus (ARC) and paraventricular nucleus (PVN) (O'Donohue *et al.*, 1985). NPY promotes fat storage, food intake, and reduces energy expenditure when it is centrally administrated to the hypothalamus, resulting in obesity (Billington *et al.*, 1991; Stanley *et al.*, 1986; Zarjevski *et al.*, 1993). Leptin deficiency causes increased expression and secretion of NPY in the hypothalamus. Ob/ob mice exhibit high levels of NPY expression in the hypothalamus that are attenuated by administration of leptin (Chua *et al.*, 1991; Stephens *et al.*, 1995).

Activation of **melanocortin 4 receptor** (MC4R) by α-melanocyte-stimulating hormone via the melanocortin-signaling pathway results in inhibition of food intake (Cone *et al.*, 1996). MC4R is a transmembrane G-protein coupled receptor that activates a cyclic adenosine monophosphate (cAMP) signal pathway when α-MSH binds. MC4R mouse knockouts exhibit hyperphagia resulting in obesity (Huszar *et al.*, 1997).

UCP2 is a mitochondrial uncoupling protein that is thought to dissipate the proton gradient across the inner mitochondrial membrane. It is homologous to UCP1, which is responsible for generating heat in brown fat by uncoupling respiration from ATP production. Unlike UCP1, UCP2 is widely expressed in white adipose tissue and skeletal muscle. There is evidence that UCP2 plays a role in eliminating reactive oxygen species (Echtay *et al.*, 2002).

CHAPTER 1: SNP DISCOVERY IN CANDIDATE GENES FOR OBESITY IN WESTERN ALASKA NATIVES

INTRODUCTION

We screened 16 exons (5043 bp) in candidate genes for obesity in 56 Western Alaska Natives by denaturing high performance liquid chromatography (DHPLC) and sequencing in hopes of finding population specific genetic risk factors for obesity. Leptin, Leptin Receptor, Melancortin 4 receptor, and Neurpeptide Y were examined because of evidence from case-control and transmission-linkage studies implementing their significant roles in weight gain (Bray *et al.*, 1999; 2000; Butler *et al.*, 1998; Chagnon *et al.*, 1997; Chagnon *et al.*, 1999; Chagnon *et al.*, 2000; Hinney *et al.*, 2003; Karvonen *et al.*, 2000; Li *et al.*, 1999; Mattevi *et al.*, 2002). Coding regions were chosen because identification of common functional SNPs were of primary interest. Based on findings by Cargill *et al.* and Stephens *et al.*, we expeced to find close to one SNP per 320 base pairs of coding DNA screened for a total of 16 SNPs, roughly half synonymous and half non-synonymous (Cargill *et al.*, 1999; Stephens *et al.*, 2001). Results by Stephens *et al.* suggest that cosmopolitan SNPs will compose 82% of those found and the remaining 18% will be either unique to the Alaska Native population or shared only with related groups (Stephens *et al.*, 2001). Of the 16 SNPs we expected to find, 3 should have been non-cosmopolitan and 13 should have been cosmopolitan.

METHODS

DNA Extraction

The DNA was extracted from 1-2 ml of whole blood with Qiagen's QIAamp Blood Midi Kit. Samples were mixed with 100 µl of Qiagen protease, 1.2 ml buffer AL, and incubated at 70° C for 10 min. They were vortexed with the addition of 1 ml of ethanol (96%) until homogenous. The lysate was applied to a QIAamp midi column and centrifuged with a Jouan CR412 centrifuge at 3000 rpm for 3 min. The filtrate was discarded, 2 ml of buffer AW1 was added to the midi column, and the samples were centrifuged at 4750 rpm for 2 min. Without removing filtrate, 2 ml of buffer AW2 was added to the midi column. The samples were centrifuged at 4750 rpm for 15 min. The midi column was removed and

placed in a clean 15 ml centrifuge tube; 200 µl of buffer AE was added to each midi column and let to incubate at room temperature for 5 min. Samples were centrifuged for 5 min. at 4750 rpm. For maximum yield, an additional 200 µl of buffer AE was added to the midi columns and equilibrated for 5 min. The samples were then centrifuged at 4750 rpm for 5 min. The samples were then stored at 4°C in 1.5 ml screw cap vials.

DNA Quantification

The PicoGreen assay was used to quantify DNA yield of the pre-anonymized samples. PicoGreen DMSO stock solution was diluted 200-fold with TE. A standard curve was made by making dilutions of lambda DNA from a 2 µg/ml stock solution. The lambda dilutions (100 µl) were mixed with 100 µl of working stock PicoGreen and incubated for 5 min. at room temperature in the dark. Lambda concentrations for making the curve were 1 µg/ml, .8 µg/ml, .6 µg/ml, .4 µg/ml, .2 µg/ml, .1 µg/ml, .01 µg/ml, and .001 µg/ml. Calf thymus DNA at 50 µg/ml, as determined by absorbance at 260 nm, was used as a positive control. From each sample, 1 µl was diluted with 99 µl of TE and 100 µl of PicoGreen working solution. Fluorescence intensity was read with Molecular Devices Analyst AD. Excel was used to fit a line through standard dilution points and the equation of the line was used to calculate the concentrations of the samples.

The concentrations of the anonymized samples were determined using the Spectra Max plate reader at 260 nm absorbance. Each well contained 10 µl of sample and 90 µl of dH_2O. Beer's Law was used to calculate the concentration of DNA based on the absorbance at 260 nm. Dilutions of calf thymus DNA were used as positive controls. Working sample concentrations of 10 ng/µl were made in microtiter plates. The diluted samples were used for all PCRs.

PCR Amplification

Three different PCR mixes were used to amplify target DNA, depending on the locus (Table 1.1). Unless noted to the contrary, PCR primer sequences were attained from Drs. Wendy Chung and Rudy Leibel at Columbia University. Touchdown PCR was used to minimize mispriming (Don *et al.*, 1991). Using Taq Gold polymerase, samples were denatured for 9 min. at 95° C, followed by 10 cycles of 95° C for 30 sec. (denaturation), 65° C for 30 sec. – 1°/cycle (annealing), and 72° C for 30 sec. (extension); Samples were then cycled 25 times at 95° C for 30 sec., 55° C for 30 sec., and 72° C for 30 sec, finishing with a 6 minute extension at 72° C. Core mix was distributed with the Qiagen Biorobot 8000 and template was added with an eight-channel pipet.

PCR products were screened for individuals heterozygous for SNPs with the Transgenomic WAVE using DHPLC. The WAVE consists of a separatory column resting within a precisely controlled oven with a UV detector attached at the end. The separation cartridge is composed of polystyrene-divinylbenzene (PS-DVB) copolymer beads. The mobile phase is a mixture of triethylammonium acetate (TEAA) and acetonitrile. TEAA is an ion-pairing reagent whose hydrophobic ethyl component interacts with the PS-DVB beads while its positively charged ammonium acetate component is attracted to the negative phosphate backbone of nucleic acids. The TEAA connects nucleic acids to the column. Acetonitrile breaks the interaction between the column the TEAA and the nucleic acid. Increasing the amount of acetonitrile serves to elute the nucleic acids.

SNP Discovery using DHPLC

SNP discovery was performed using partially denaturing conditions under the mode "mutation detection." The average temperature range for partially denaturing conditions was 52-75° C. In this temperature range, fragments are separated on the basis of size and sequence. SNP discovery is dependent on the formation and separation of heteroduplexes from homoduplexes. Heteroduplexes are formed by the mispairing of parent strands.

Table 1.1 PCR Mix Concentrations derived from Ruth Stafford (R), Dr. Kevin McCracken (DM), and Dr. McCracken's difficult (DMD) PCR mix.

Reaction components	R concentrations	DM concentrations	DMD concentrations
dH_2O	–	–	–
10 X PCR Gold reaction Buffer	1X	1X	1X
2.5 mM dNTPs	.1 mM	.25	.3
25 mM $MgCl_2$	1.75 mM	2.5 mM	3.0 mM
5 U/µl Taq Gold	1.56 U/50 µl reaction	1.25 U/50 µl reaction	1.5 U/ 50 µl reaction
10 µM forward primer	.4 µM	.5 µM	.6 µM
10 µM reverse primer	.4 µM	.5 µM	.6 µM
10 ng/µl Template DNA	15-30 ng	15-30 ng	15-30 ng

Mispairing of parent strands may be induced by mixing samples with a wild type or if the sample being analyzed is at least diploid and heterozygous. Heteroduplex formation was facilitated by denaturing PCR products at 95°C for five min. and lowering the temperature to 25° C at a rate of 0.1° C/4 sec. (Schmitt *et al.*).

The sequence of the PCR product was entered into the WAVE's Navigator software to generate predicted fragment melt profiles based on temperature. The temperature in the program was adjusted to make the amplicon have a helical fraction between 70% and 85%, the range at which SNPs are optimally resolved (Schmitt *et al.*). Although, SNPs should be detected when the helical fraction is between 30% and 98%, as long as greater than half the fragment has a helical content above 50%. The predicted melt profile generated by the Navigator software gives the user an idea of what temperatures might detect SNPs in what region of the fragment but temperatures were still empirically optimized based around what temperatures appeared most promising in the melt profile. Fifteen exons in four candidate obesity genes were screened in 56 Alaska Native samples with DHPLC (Table 1.2). Exon 2 of MC4R was screened by sequencing because it was double the length of the suggested optimal amplicon screening size and contained many different melting domains. Experience with DHPLC screening of the similarly sized MC4R exon 1 prompted the change in screening method for MC4R exon 2. When a heteroduplex was detected two homozygotes and two heterozygotes were sequenced to characterize the SNP responsible for the heteroduplex.

PCR products were cleaned for sequencing reactions using the Qiaquick PCR purification kit or with the Qiagen Biorobot 8000 along with the Qiaquick Multiwell PCR Purification kit (Qiagen, 2002; 2002). For the cycle sequence reaction, 2 µl of Big dye, 3 µl of purified PCR product, 0.3 µl of 10 mM primer, and 4.7 µl of dH_2O were added for a total reaction volume of 10 µl. The cycle sequencing protocol was 96° C for 1 min. followed by 24 cycles at 96° C for 10 sec., 50° C for 5 sec., and 60° for 4 min.

Table 1.2 Obesity Candidate Gene Exons Screened for SNPs by DHPLC and Sequencing

Amplicon	Forward Primer	Reverse Primer	Length of Amplicon	DHPLC SNP Screen Temp. in Celsius	Optimal DHPLC Temp. in Celsius	Sequence Verification of SNP
LEP ex 2	AAGAAGCCCATCCTGGGA	GCTGGCTGCAGTTCTACTTTG	209	60, 60.2, 60.5		
LEP ex 3a	CCACATGCTGAGCACTTGTT	TGGCAGCTCTTAGAGAAGGC	250	61, 61.3		
LEP ex 3b	TGGAGAACCTCCGGGATCT	GTTCCTTCCCTTAACGTAGTCC	251	61.6, 64.6		
LEPR ex 1	TCCACAGACAACTTATATATATGTG	GCTAAAGGCAAAATCTCTACCATG	193	55.2, 53.7, 56.7, 57.2		
LEPR ex 2a	TCAGATACTTTCTATTCATGTCTTAGT	ATAAGTTAGAAAAGTGAGTACCACTTG	279	52.5, 56, 57		
LEPR ex 2b	AATTCGAATGGACATTATGA	GTATCAAAGAATTAAAAAACATTGTTC	252	52.4, 54.4, 57.1		
LEPR ex 3	TTTTTTTTGTTTTTTTTTTTTTTTAAATTC	TAAAAAAACTGTGTATTAGAAATGC	205	54		
LEPR ex 4	TCCTGCTTTAAAAGCCTAATCCAGTATTT	AGCTAGCAAATATTTTTGTAAGCAAT	367	52.5, 57, 58	58	YES
LEPR ex 5	GACTTTATTTTATTCAGCTATAATTGT	GCAGAGGGTAATTGCTATGGGAC	300	52.8, 54.8, 55.8	54.8	NO
LEPR ex 7a	GGCAGTGTAACTCTGGAAGTGTTGT	AGTACACTGCATCATAGGTAAACT	343	59.9		
LEPR ex 7b	ATCTATAAGAAGGAAAACAAGATTGT	GGACTCTAATATAAGGAGGGTCCAT	305	56.8		
NPY ex 1	GCGCGTGGGTGCTCTGAATC	GGATCTCCTGGTGTGCAGGCAC	301	66.2, 63.2, 64.8	66.2	YES
NPY ex 2	TGCTTCATACACCTAGCTTG	TTCATCCTCATTTCTGCCGA	364	55.4, 59.5, 61.5	59.5	YES
NPY ex 3	CTTCAGATCTAAATGTCTCAC	AGTGGCTGCATGCATTGGTAG	188	59, 59.4, 59.8, 60		
MC4R ex 1	ATCAATTCAGGGGGACACTG	GACAGCACTACTATCTGAGT	614	57.8, 58.3, 60.9		
MC4R ex 2	ATGCTCTCCAGTACCATAACA	TGCAGAAGTACAATATTCAGG	622			

Samples were cleaned via a Millipore multiscreen filter plate sephadex G-50 spin column (Millipore, 2004). A 96 well aluminum block was filled with sephadex G-50 and a plastic scraper was used to level the wells. The Millipore filter plate was placed upside down on the block and then flipped to fill the sephadex into the plate. Water (300 μl) was added to each well and the plate was let to sit for 2-3 hours. The plate was then spun at 1000 RCF for 5 min. An additional 150 μl of water was added to the plate and was spun again at 1000 RCF for 5 min. Cycle sequencing products were added to the columns and the plate was spun for 5 min. at 1000 RCF. The collection plate containing the purified products was then incubated for 15 min. at 90° C to evaporate the water. Samples were stored at –20° C in the dark until they were ran on an Applied Biosystem 3700 Automated Sequencer. Prior to reading, 15 μl of formamide was added to each well, and the samples were denatured at 95° C for 5 min. and then placed on ice.

After sequencing, the SNPs were looked up in dbSNP and characterized using a NCBI index. NCBI indexes SNPs in dbSNP by two different accession numbers. The "rs[NCBI SNP ID]" accession number refers to the NCBI reference SNP cluster identifier, and it is attached to all records associated with the locus. The "ss[Assay Id]" refers to the individual submission record. We used the "rs" accession number because it is inclusive of all records associated with the site.

Phenotype

Body-mass index (BMI) is equal to the weight of an individual in kg divided by the square of their height in meters. BMI was used to divide samples into three categories: lean (BMI $\leq$ 23), in-between (23< BMI <30), and obese (BMI $\geq$30). From each population, 13 lean, 41 in-between, and 21 obese were selected for SNP screening of candidate obesity gene exons. Samples had to have a minimum of 10 μg of DNA for use in the study. Because they were not originally collected for genetic research, samples were de-identified on individual and population levels in accordance with the University of Alaska Fairbanks Institutional Review

Board's criteria for use of the samples for human genetic research. Chosen samples were transferred to 4 unmarked boxes containing 75, screw-top, 1.5 ml, skirted tubes labeled "L" for lean, "I" for in-between, and "O" for obese. Two people not associated with the study mixed the tubes within each box. The box tops were then labeled A, B, C and D. Two people not associated with the study de-identified the boxes. The tubes were then labeled with a number, corresponding to their de-identified box top. In this procedure, all identifiers were removed (except for BMI class) and samples were anonymized according to individual and village.

RESULTS

A total of 5043 base pairs were screened. Out of 16 amplicons screened, SNPs were detected in LEPR exons 4 and 5 and NPY exons 1 and 2 (Figure 1.1). The DHPLC detected SNP in LEPR exon 5 was the only SNP that sequencing did not verify. The sequences of the other SNPs showed that they were not newly discovered or exclusive to the Alaska Native population: NPY exon 1 rs5573 (A/G), NPY exon 2 rs5574 (C/T), and LEPR exon 4 Glu223Arg rs1137101 (A/G). None of the NPY SNPs cause an amino acid change.

DISCUSSION

Comparison of SNP finds with other studies

Including the non-sequence verified SNP in LEPR exon 5, 1 SNP per 1.3 kb of coding region DNA screened were found. This is considerably less than might be expected, based on other studies. Stephens *et al.* found 1 SNP per 294 bp of coding region screened when they studied 313 human genes in 82 ethnically diverse individuals (Stephens *et al.*, 2001). The results of Haluska *et al.* were similar to Stephens *et al.* in that they found 1 SNP per 225 bp of coding region screened when they examined 75 candidate genes for blood pressure homeostasis in 74 humans of African and Northern European descent (Halushka *et al.*, 1999). Cargill *et al.* reported a ratio of 1 SNP per 346 bp coding DNA screened when

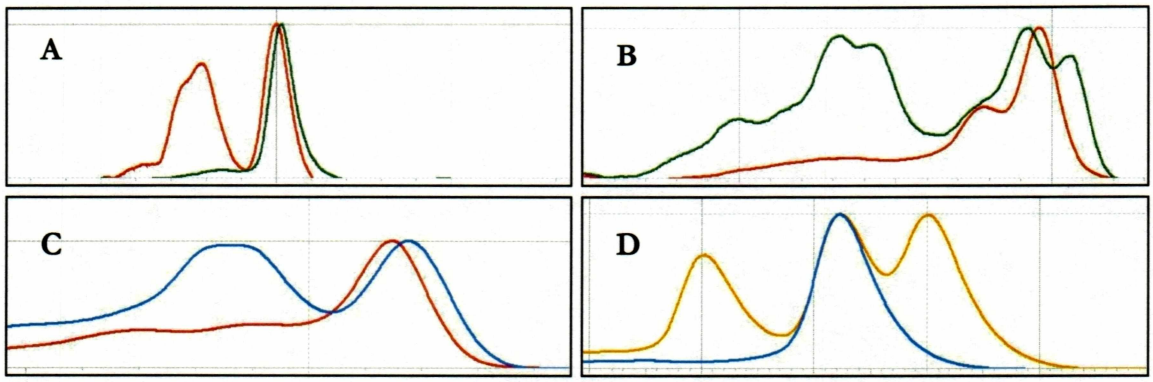

Figure 1.1 DHPLC chromatographs for LEPR ex4 SNP (A), LEPR ex5 SNP (B), NPY ex1SNP (C), and NPY ex2 (D). The heteroduplex is the far left peak(s). Elution window width: 2-4 min. for A, 2-4.5 min. for B, 4-6 min. for C, and 2.3-2.7 min for D.

they studied 106 genes associated with cardiovascular disease, endocrinology and neuropsychiatry in 57 individuals of European, African-American, African Pygmie, and Asian descent (Cargill *et al.*, 1999).

Reasons for lower frequency of SNP finds

Why did we find a lower frequency of SNPs in the coding regions of genes examined? The answer lies in sampling only one ethnic group, screening fewer genes (bps), and missing SNPs. The majority of human SNPs are cosmopolitan but by studying one population, W. Alaska Native population, we excluded the possibility for detection of SNPs specific to other populations (Stephens *et al.*, 2001). Sampling a small number of base pairs made it more likely that the ratio of SNPs detected to base pairs screened was a fluctuation away from the expected. We expect that screening a greater number of base pairs would cause the SNP frequency to approach those found by Haluska *et al.*, Cargill *et al.*, and Stephens *et al.* Non-detection of SNPs could have occurred because they are rare, fixed, or missed due to methodological screening error. It is likely that some of the SNPs are rare or fixed due to the reduced genetic diversity in the Alaska Native population (Budowle *et al.*, 2002; Shields *et al.*, 1993). The DHPLC screening methods were partially tested by sequencing 14 individuals for 4 amplicons that showed no heteroduplexes. The sequence data agreed with the DHPLC, showing no individuals heterozygous for SNPs. Since all amplicons negative for heteroduplexes were not congruently screened with sequencing we are not able to definitively say whether the DHPLC screen method did or did not miss SNPs. However, based on reports of SNP frequencies in human exons and SNP reports specific to the exons screened, it is strongly suspected that not all SNPs were detected.

Expected false positive and false negative rates for DHPLC

Cargill *et al.* used DHPLC and variant detection arrays (VDAs) to screen 106 genes and confirmed SNP discoveries with sequencing. The VDAs had a false positive screening rate of 45%, while DHPLC had a rate of 40%. Our false-positive rate was 25% for the DHPLC

screen. To determine the false-negative rate, Cargill *et al.* sequenced 10 genes in 20 individuals; VDA detected 34/40 SNPs found by sequencing and DHPLC discovered 46/53 SNPs identified by sequencing. The reported false-negative rate for DHPLC is not great enough to explain the difference between the SNP discovery rate in our study and those reported by Haluska *et al.*, Cargill *et al.*, and Stephens *et al.* but it could describe a portion of the difference.

Examination of exons lacking SNP finds

Exons that did not show any SNPs in the screened Alaska Native population were examined in dbSNP. Several SNPs reported in dbSNP were not found in the Alaska Native samples. Two SNPs in LEP exons 2 and 3, rs1800583 and rs3750043, were reported without any frequency data. No functional SNPs were found in Leptin exons 2 and 3 matching results by Li *et al.* when they studied LEP coding regions in obese women (Li *et al.*, 1999). LEPR exon 2 rs1137099 was reported in dbSNP without any frequency data, and LEPR exon 7 rs1805134 showed a minor allele frequency of .272 in dbSNP; Celera showed the minor allele frequency to be .26 for Caucasians and .31 for African Americans. NCBI's dbSNP showed three SNPs for MC4R: rs2282556, rs229616, and rs1016862. Rs2229616 had a minor allele frequency of .026 in a multinational screening group. NPY had two SNPs, rs5575 and rs5576, in exon 3 that were not detected in the Alaska Native screen population. The minor allele frequencies were 0.1 and .05 respectively. For the dbSNPs lacking frequency data we cannot say whether we would expect to find them in the Alaska Native population. For the SNPs having reported minor allele frequencies greater than .10 we would be more apt to expect them to exist in the Alaska Native population, since common SNPs are usually ancient and thus cosmopolitan (Risch, 2000).

CONCLUSION

In conclusion, the frequency of SNPs per base pair is likely to increase as more candidate gene exons are screened. Screening more base pairs will help to overcome the stochasticity

of sampling effects. Future SNP discovery studies will benefit from pooling Alaska Native DNA with Caucasian or African DNA to increase the chance of detecting SNPs fixed in the population. Since common SNPs are of more interest, it is not necessary to increase the number of individuals screened. SNPs > 10% frequency are more likely to be cosmopolitan and have results significant to other populations, providing support for causality of the allele (Risch, 2000).

Call for a change in methodology

The SNP screen results call for a change in methodology. SNP screening can be costly, time consuming, and inefficient when the results are redundant. Currently, dbSNP contains 3.7 million SNPs, providing plenty of options for the design of candidate gene case-control association studies (Jiang *et al.*, 2003). A more efficient study design would use previously characterized common functional variants to test candidate genes for association with the variation of a particular phenotype. Non-cosmopolitan SNPs would likely be missed by this method but that is of no consequence because they are likely to be rare SNPs arising from recent mutations. Use of common SNPs would be beneficial because study results could be compared to other populations. Replication of results in different populations would help support the causality of the locus.

CHAPTER 2: TESTING 3 NPY SNPS FOR ASSOCIATION WITH BMI IN WESTERN ALASKA NATIVES

INTRODUCTION

Neuropeptide Y (NPY) has been identified as a key player in obesity pathophysiology. Lack of NPY in ob/ob-NPY-/NPY- double knockouts attenuates the obesity syndrome observed in ob/ob single knockouts (Erickson *et al.*, 1996). Intercranial infusion of NPY increases feeding and reduces metabolic rate in rats (Kotz *et al.*, 1998; Stanley *et al.*, 1986; Stanley and Leibowitz, 1984). Familial linkage and case-control studies have further implemented NPY as a candidate gene for obesity in human populations (Bray *et al.*, 1999; 2000; Karvonen *et al.*, 2000).

Bray *et al.* used sibling pair linkage analysis for microsattelites close to NPY to test for obesity linkage in 59 Mexican-American families (170 obese sibling pairs out of 545 total pairs); they found significant linkage within sibling pairs (Bray *et al.*, 1999). Bray *et al.* then investigated polymorphisms in the promoter and coding regions of NPY by sequencing, finding a significant association between -880I/D and body fat patterning in a non-obese (BMI less than 30) subsample of Mexican-Americans (Bray *et al.*, 2000). Evidence of familial linkage along with the reported significant association of the –880I/D variant facilitated the decision to investigate NPY variants in the Alaska Native population.

In this case-control study of the Western Alaska Native population, we tested for association of 3 NPY SNPs with BMI, with the hypothesis that the SNPs are associated with BMI category by being in linkage disequilibrium (LD) with a causative allele. Linkage disequilibrium is the nonrandom association of genetic markers in a population. Our aim was to determine whether any or all-3 NPY SNPs genotyped may be informative for risk of obesity in the Western Alaska Native population.

METHODS

Three NPY SNPs were genotyped with three different methods because some methods worked better than others, depending on the locus. SBEX was used to genotype the population for rs5573, a synonymous SNP in the first exon of NPY. DHPLC was used to genotype rs5574, a synonymous SNP in the second exon of NPY. Finally, ABI's Assay on Demand C 11164478_20 was used to genotype NPY rs16478, a SNP 5' to the transcription start site of NPY. The program Genepop was used to calculate the Hardy-Weinberg exact test and Fwc(st).

SBEX uses fluorescently labeled dye terminators to extend a primer by a single base, the SNP of interest. The dye is either R110 or Tamara. Fluorescence polarization (FP) is used to discriminate between incorporated and non-incorporated nucleotide fluorescence. Molecules are excited using plane-polarized light. If a molecule is excited by plane-polarized light then the light it emits is also polarized (Perrin, 1926). FP is proportional to molecular volume and weight; larger molecules emit greater FP than smaller molecules. A dye terminator attached to the end of a primer emits a greater amount of polarized light than an unattached dye terminator (Figure 2.1). FP is independent of fluorescence intensity because it measures the ratio of polarized light in the vertical and horizontal fields.

Samples were amplified via PCR prior to the SBEX using Perkin Elmer's suggested PCR concentrations (Table 2.1). Touchdown PCR conditions were used, as specified in chapter 1. A SBEX primer (forward 5'-GGAGGACATGGCCAGATACTACTC-3') adjacent to rs5573 was ordered from Qiagen. Perkin Elmer's AcycloPrime-FP A/G SNP detection kit (ACP101A) was used for the SBEX. PCR products were purified using 2 μl of Perkin Elmer's 1X PCR clean-up reagent. The clean-up reagent was added to 5 μl of PCR product and incubated in a Tetrad DNA engine at 37° C for 60 min. followed by 15 min. at 80° C, inactivating the enzyme. The cleaned PCR product was added to 13 μl of core SBEX mix in a black MJ Hardshell skirted plate and cycled at 95° for 2 min followed by 50 cycles of 95° for 15 sec and 55° for 30 sec. (Table 2.2).

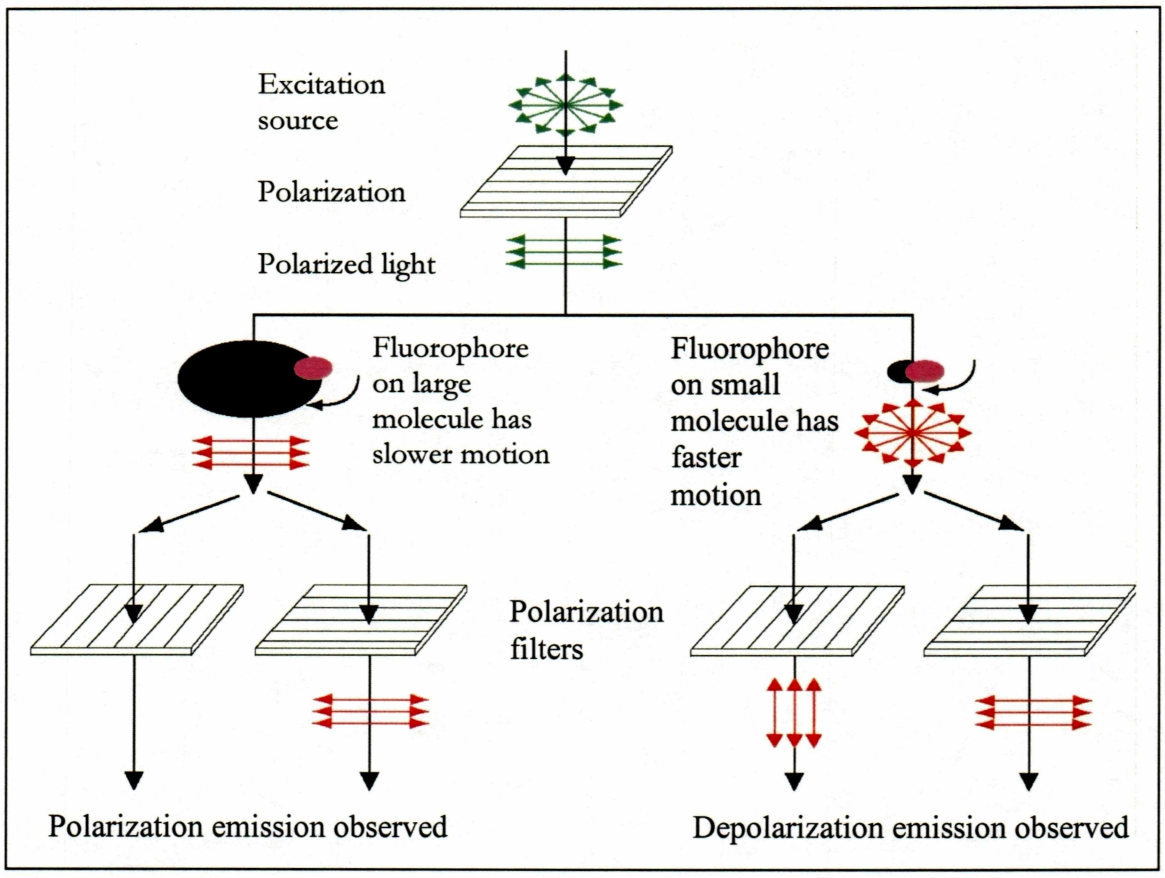

Figure 2.1 Principle of fluorescence polarization (Chen *et al.*, 1999).

Table 2.1 Perkin Elmer PCR Conditions for Samples

Reaction components	µl per reaction	[Final]
H_2O	12.475	
10X PCR Gold buffer	2	1
10 mM dNTPs	0.2	100
25 mM $MgCl_2$	1.4	1.75
5 U/µl AmpliTaq Gold	0.125	0.625 units
10 µM forward primer	0.4	0.2
10 µM reverse primer	0.4	0.2
10 ng/µl template DNA	3	15-25ng
Total	20	

Table 2.2 SBEX Core Mix

Reagent	µl/rxn
Acyclopol	.05
10X RxN Buffer	2
Acyclo terminator mix	1
SNP primer (10 uM)	0.5
Water	9.45
Total volume	13

The Molecular Devices Analyst AD was used to read polarized fluorescence of the SBEX reaction. Samples were first read with the R110 filter (ex. 490-10 and em. 520-10) and then read with Tamara (ex. 550-10 and em. 580-10). Both reads were saved to the same file so they could be opened with the program Allele Caller. Allele Caller creates a visual representation of the FP values for each sample. Each axis represents FP associated with a particular base.

The WAVE was used to genotype NPY exon 2 rs5574. Buffer A was 5% .1M TEAA and buffer B is 5% .1M TEAA and 25% acetonitrile. The WAVE was set for mutation detection with the oven temperature set at 59.5° C. The gradient (%B per min.) was set at 2 and started at 56.3% B and ended at 65.3% B.

An Assay on Demand, C_11164478_20, was used to genotype 58 samples from population B for a SNP (rs16478) 5' of the transcription start site of NPY. Each reaction had a core mix consisting of 9.25 µl of H_2O, 12.5 µl of 2X Taqman PCR master mix, and 1.25 µl of SNP genotyping assay mix. The Qiagen Biorobot 8000 was used to distribute 23 µl of core mix to each well and 2 µl of Template was added manually. Samples were incubated at 95° C for 10 min. and then cycled 40 times at 92° C for 15 sec. followed by 60° C for 1 min. The samples were then covered with adhesive foil and sent to Applied Biosystems. They were read as an endpoint assay on the Taqman.

RESULTS

The SBEX reaction for NPY exon1 rs5573 yielded distinct genotype groupings for all populations: A, B, C, and D (Figure 2.2). The genotypes were totaled and divided according to their respective BMI group (Table 2.3). Genepop was used to test the locus for Hardy-Weinberg equilibrium and to calculate Fwc(st) (Table 2.4). Fst is a F-statistic developed by Sewall Wright to measure the amount of population subdivision and the reduction of heterozygosity. It compares the heterozygosity expected in a total population to the

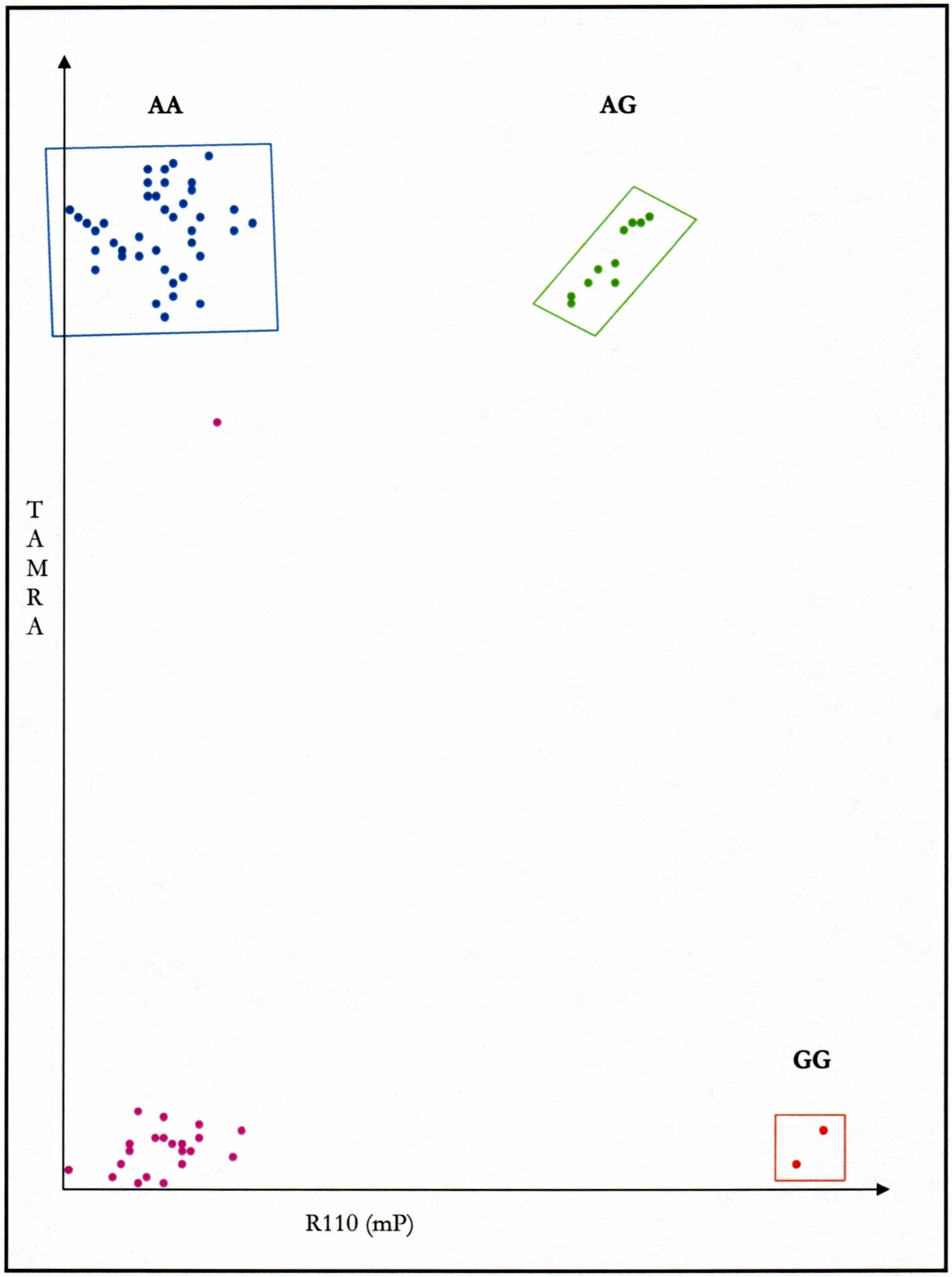

Figure 2.2 SBEX genotyping of population A for NPY rs5573.

Table 2.3 NPY Exon 1 A/G SNP, rs5573, Genotype Frequencies Determined with SBEX.

Phenotype	AA	AG	GG	Totals
Lean	42	4	2	48
In-between	103	30	3	136
Obese	60	8	1	69
Totals	205	42	6	253

Table 2.4 Hardy-Weinberg Test and Fwc(st) for NPY SNPs.

SNP	Degrees of freedom	$\chi 2$ statistic for Hardy-Weinberg Probability test	Probability	Fwc(st)
rs16478 promoter	2	--	.27 (ns)	--
rs5573 ex 1	8	9.54	.30 (ns)	.0012
rs5574 ex 2	8	7.4	.49 (ns)	.0240

mean heterozygosity across all subpopulations. Low levels of gene flow should yield higher Fst values. The probability of fit as being worse by chance was >.05; so, the Hardy-Weinberg model cannot be rejected. The Fwc(st) statistic was <.05, showing little population differentiation between populations A, B, C, and D, with respect to the locus (Wright, 1978).

Associations between rs5573 and obesity were tested under additive and non-additive models using χ^2-distributed test statistics. A trend test was used to model the effect of an additive allele, and both a genotype case-control statistic and a G-test of independence were used to model nonadditive effects (Table 2.5). A trend test compares the increment of the likelihood of having a phenotype between heterozygotes and homozygotes. If the likelihood of having the phenotype is significantly similar then the hypothesis that the locus is associated with the trait in an additive manner is supported. The G-test of independence tests the goodness of fit of observed and expected values (Sokal and Rohlf, 1995). G-tests have a higher level of type I errors (rejection of a true null hypothesis); so, William's correction factor is applied to better model the chi-square distribution (Williams, 1976). We cannot reject the null hypothesis that the genotype at rs5573 is independent of BMI grouping. The minor allele frequency of 0.11 was much lower than those reported for other ethnic groups (Table 2.6).

The WAVE yielded distinct genotypes for NPY exon 2 rs5574 that were tallied and grouped according to BMI group (Figure 2.3 and Table 2.7). Genepop was used to test whether the population was in Hardy-Weinberg equilibrium for rs5574. The chi-square test resulted in a P-value >.05; so, the hypothesis that the population was in Hardy-Weinberg equilibrium could not be rejected (Table 2.4). Fwc(st), calculated using Genepop, was .024, showing little genetic differentiation between subpopulations (Table 2.4). Tests for independence of BMI group from rs5574 genotype were not significant for both additive and nonadditive models. The minor allele frequency, .44, is close to the .376 reported by dbSNP in a multi-ethnic sample set (Table 2.6).

Table 2.5 χ2-distributed Test Statistics for Independence of Genotype from Phenotype in NPY SNPs.

SNP	Sample size	Trend test for additive allele using only Obese and Lean	Non-additive χ^2 statistic (4df)	G-test with William's correction (4 df)
rs16478 promoter	37	5.9*10^-5 (ns)	4.2 (ns)	4.06 (ns)
rs5573 ex 1	253	1.3*10^-3 (ns)	7.39 (ns)	7.09 (ns)
rs5574 ex 2	271	1.1*10^-5 (ns)	1.56 (ns)	1.53 (ns)

Table 2.6 NPY Minor allele frequencies.

	Caucasian (Celera)	Chinese (Celera)	Japanese (Celera)	African American (Celera)	Multi-ethnic (dbSNP)	Alaska Native
Promoter rs16478	.29	--	--	.36	--	.09
Exon 1 rs5573	.5	.44	.38	.32	.317	.11
Exon 2 rs5574	--	--	--	--	.376	.44

Table 2.7 NPY Exon 2 C/T SNP, rs5574, Genotype Frequencies Determined with DHPLC.

Phenotype	CC	CT	TT	Totals
Lean	14	15	19	48
In-between	45	36	67	148
Obese	25	21	29	75
Totals	84	72	115	271

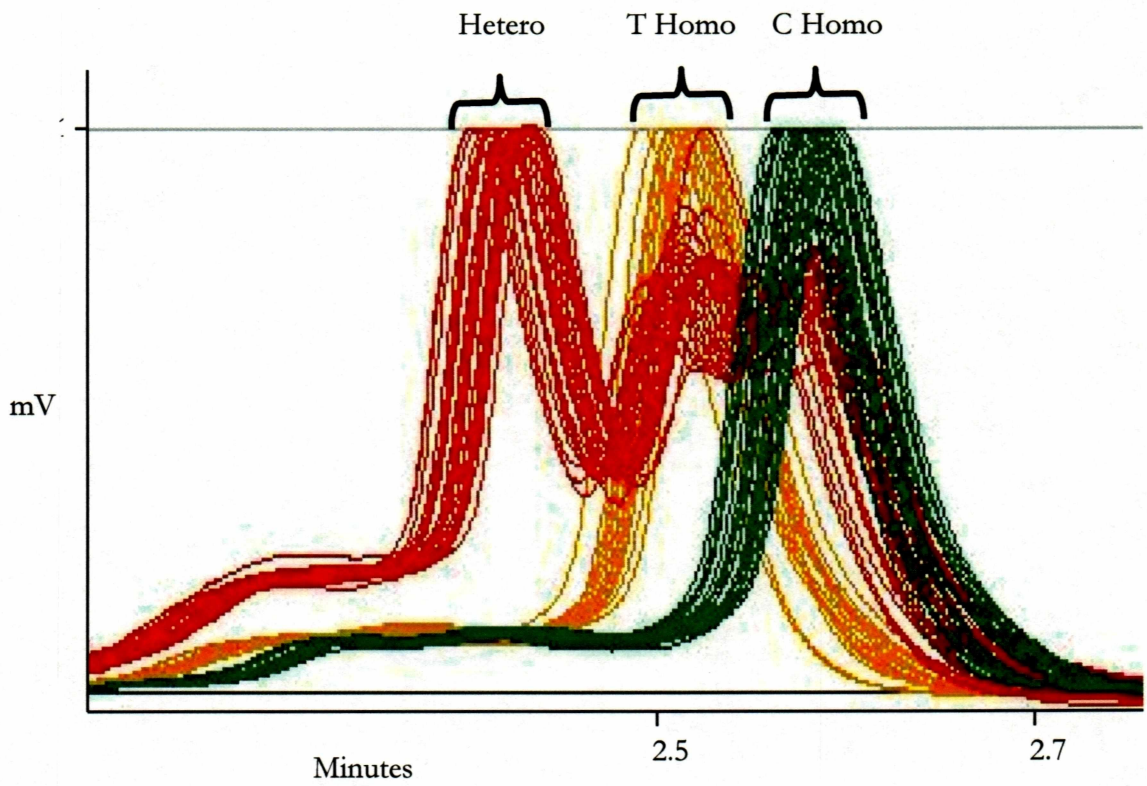

Figure 2.3 WAVE Chromatogram of the Separation of NPY ex2 SNP Heteroduplex and Homoduplexes.

Figure 2.4 displays results for the genotyping of NPY promoter SNP, rs16478, and Table 2.8 shows the genotype counts for each BMI group. The P-value for the Hardy-Weinberg test was >.05; so, the hypothesis that the population is in Hardy-Weinberg may not be rejected. The hypothesis that the NPY promoter SNP is independent of BMI category could not be rejected for both additive and non-additive models (Table 2.5). The minor allele, A, frequency was .09, which was much less than other ethnic groups (Table 2.6).

DISCUSSION

Hardy-Weinberg could not be rejected, lending support to the idea that the populations are randomly mating. One study showed that villages used to be smaller with much inter-village movement; they were located within 62 miles of two other villages and roughly half of mothers and fathers did not live in their birthplace (Scott and Wright, 1983). Genetic data shows that the Alaska Native population has lower genetic diversity when compared to Caucasians and Africans, and they have little genetic differentiation among subpopulations (Budowle *et al.*, 2002; Shields *et al.*, 1993). Our data support previous findings with low Fwc(st) values. Although mating is unlikely to be totally random, the demographic and genetic data suggest that there is enough homogeneity between the Alaska Native villages studied to combine allele frequency data from the 4 populations for the tests of independence of SNP genotype from BMI group.

The minor allele frequencies for the NPY exon 1 and promoter SNPs were very different from those reported for other populations. It is interesting that the SNP in exon 2 was the only NPY SNP to have a minor allele frequency close to a reported value in another sample set. The NPY exon 2 SNP was also unique in that it had the highest Hardy-Weinberg probability. This raises the question of why the promoter and exon 1 SNPs have lower Hardy-Weinberg probabilities. It would be tempting to hypothesize that they are in LD with a locus affected by evolutionary forces not strong enough to be detected by the chi-square test for Hardy-Weinberg.

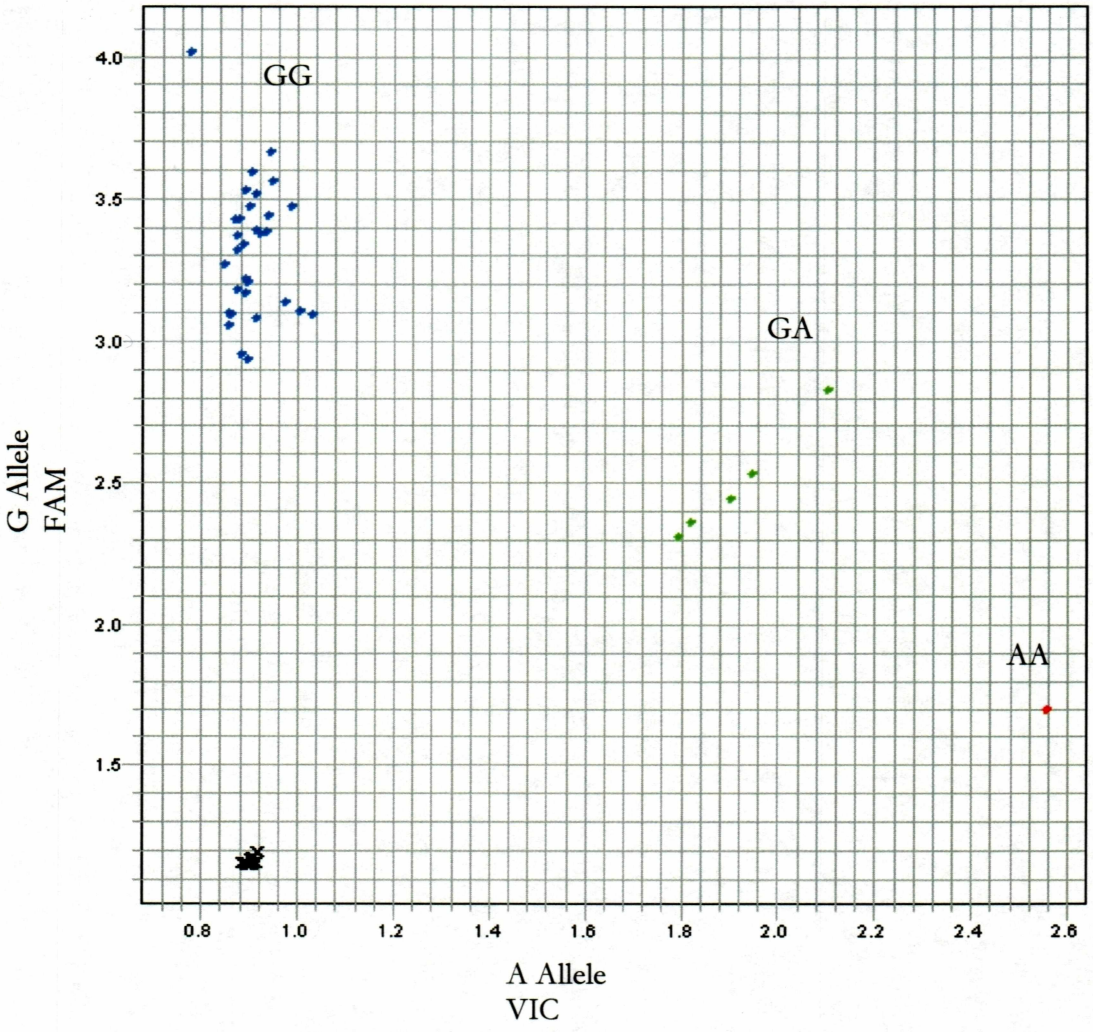

Figure 2.4 Allelic Discrimination Plot for a NPY Promoter SNP Generated with Applied Biosystems Taqman.

Table 2.8 NPY Promoter A/G SNP, rs16478, Genotype Frequencies Determined with Taqman.

Phenotype	AA	AG	GG	Totals
Lean	0	0	11	11
In-between	1	4	13	18
Obese	0	1	7	8
Totals	1	5	31	37

The null hypothesis that the NPY SNP genotype was independent of BMI group could not be rejected because there was either no effect of the SNP on BMI or there was not enough power to detect an association. It is likely that the NPY SNPs examined had no direct functional effect on BMI because they consisted of non-coding and synonymous changes, but it is plausible that they could be in LD with functional SNPs. If these SNPs were in LD with a functional SNP affecting obesity, it would be difficult to detect an association because variables known to affect BMI could not be corrected for due to the anonymization. In addition, if NPY carries polymorphisms of small effect, accounting for ~5% of trait variation, an association is not likely to be detected because our sample size was less than 500 (Long and Langley, 1999; Palmer and Cookson, 2001). Therefore, to best increase the probability of detecting a SNP/phenotype association we must increase our sample size above 500 individuals and keep all descriptors that may affect weight so the appropriate statistical correction may be applied. Since we do not have any linkage disequilibrium data on the loci of interest in the Western Alaska Native population it would be best to choose common SNPs, having allele frequency data in at least one population, that alter an amino acid or potentially change gene expression.

Having allele frequency data indicating that a SNP is common would give confidence to the decision to use the SNP in an Alaska Native study. Common SNPs in one or more populations are more likely to be common in other populations. Even though this set of Alaska Native DNA samples has indicated that some of their SNP frequencies are different from other study populations, all of the NPY SNPs have minor allele frequencies large enough to be considered common. A benefit of using common SNPs is the ability to compare results between different populations.

CHAPTER 3: TESTING UCP2 –866 A/G, rs659366, FOR ASSOCIATION WITH BMI IN WESTERN ALASKA NATIVES

INTRODUCTION

UCP2 is widely expressed and codes for a mitochondrial uncoupling protein (Gimeno *et al.*, 1997; Jaburek *et al.*, 1999). UCP2's role in energy expenditure and heat production and its close proximity to microsatalite markers linked to resting energy expenditure in humans provides support of the classification of UCP2 as a candidate obesity gene (Boss *et al.*, 2000; Bouchard *et al.*, 1997). Evans *et al.* found a positive association of a 45 bp insertion in UCP2 and morbid obesity, and Esterbauer *et al.*. detected an association of a common UCP2 promoter region SNP, rs659366, with decreased risk of obesity in humans of Bavarian and Austrian German descent, increased mRNA expression, and increased reporter gene expression (Esterbauer *et al.*, 2001; Evans *et al.*, 2000; Evans *et al.*, 2001).

We wanted to test for the –866, rs659366, association with BMI category in the W. Alaska Native population. Detection of the association would provide support to the hypothesis that the locus describes a portion of BMI variation in human populations. Non-rejection of the null hypothesis could imply that the association is population specific, Esterbauer *et al.* study had erroneous results due to cryptic underlying population substructure, or that our study design lacked the power to detect the association.

METHODS

UCP2 rs659366 was genotyped using SBEX and the Molecular Devices Analyst AD. PCR cycling conditions and R mix concentrations were used, as described in chapter 1. PCR primers were designed using the Saccharamyces Genome Database web primer design program; the forward primer sequence is 5'-GGGAATCCGTTTCCTCATTGT-3', the reverse primer is 5'- TGAGAAAAGGCGTCAGGAGAT-3', and the SBEX forward primer

is 5'- TGAGAAAAGGCGTCAGGAGAT-3' (Dolinski *et al.*, 2003). The SBEX reaction was carried out according to methods described in chapter 2.

RESULTS

The SBEX reaction of UCP2 –866 yielded distinct genotype groupings that were then tallied according to BMI group (Figure 3.1 and Table 3.1). The hypothesis that the population was in Hardy-Weinberg equilibrium could not be rejected because the Hardy-Weinberg probability test yielded a P-value >.05 (Table 3.2). The Fwc (st) indicated that there was little genetic variation between the four villages, with a value < .05 (Table 3.2). The minor allele is "G" and the frequency was .39. A multi-ethnic sample set from dbSNP showed "A" to be the minor allele with a frequency of .333. All tests for independence for both additive and non-additive models were not significant (Table 3.3). The null hypothesis that the genotype at UCP2 –866 is independent of BMI group could not be rejected.

DISCUSSION

Previous findings show that the UCP2 –866 A allele increases transcription of UCP2, is associated with decreased obesity risk, and is associated with increased glucose oxidation and decreased lipid oxidation in European populations (Esterbauer *et al.*, 2001; Krempler *et al.*, 2002; Le Fur *et al.*, 2004). Our results differ, in that they do not show an association between the –866 variant and BMI. We may not have been able to detect an association of –866 and BMI because of small sample size, the inability to correct for variables affecting BMI, and a weak genetic effect of the locus (Hirschhorn *et al.*, 2002). Alternatively, the associations of the –866 variant with obesity and increased lipid oxidation may only exist in the European population, from which they were reported. An area of concern with Esterbauer *et al.*, study is that there was no control for possible underlying population stratification. Populations were said to be mostly of Barvarian and Austrian descent but the idea that their results could be confounded by genetic admixture was not addressed. Spurious allelic association has been implicated in studies such as Knowler *et al.*, study on the

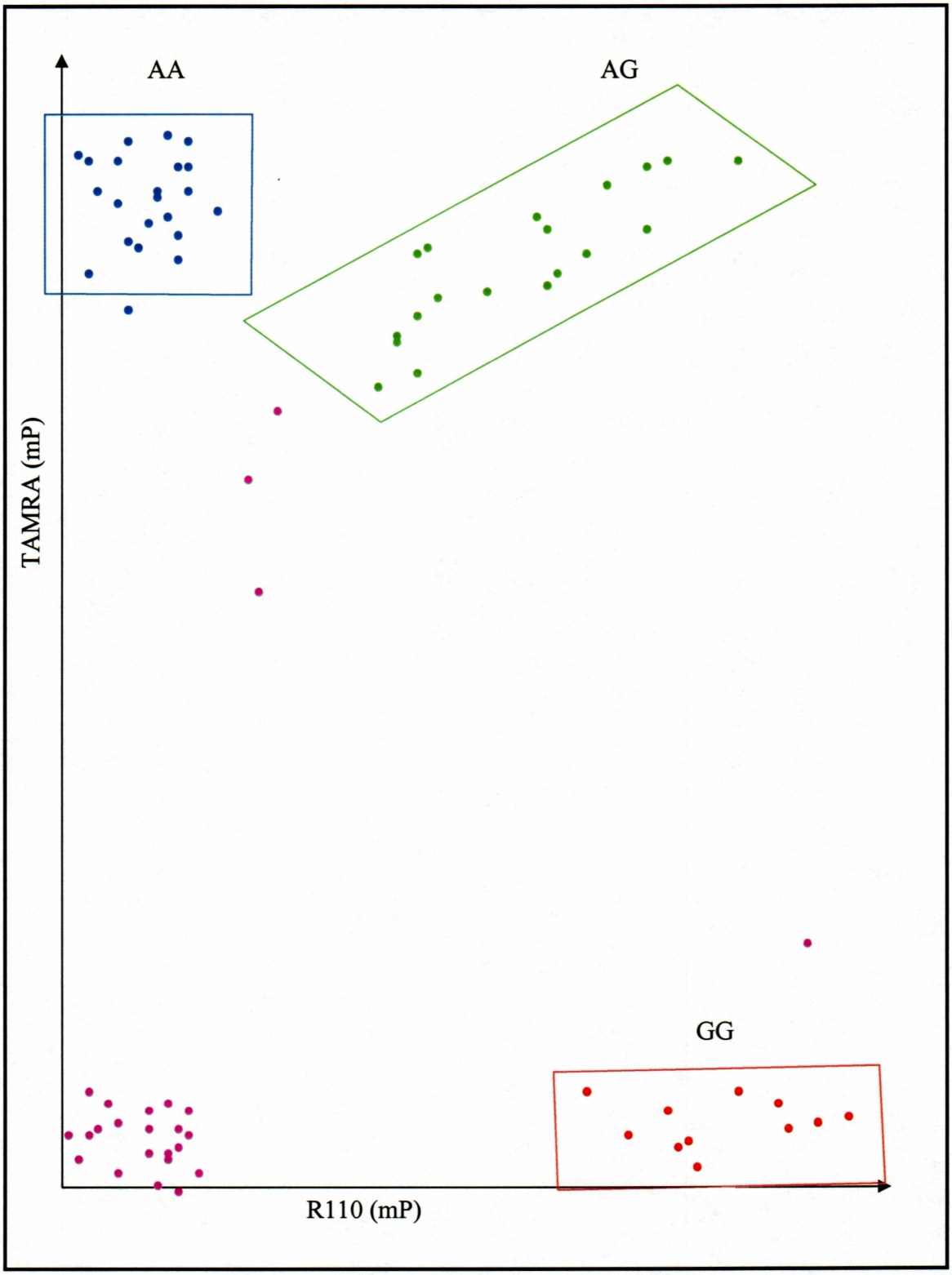

Figure 3.1 SBEX genotyping of population B for UCP2 -866.

Table 3.1 UCP2 Promoter A/G SNP, rs659366, Genotype Frequencies Determined with SBEX.

Phenotype	AA	AG	GG	Totals
Lean	16	27	7	50
In-between	51	65	18	134
Obese	22	28	12	62
Totals	89	120	37	246

Table 3.2 Hardy-Weinberg Test and Fwc(st) for UCP2 -866.

SNP	Degrees of freedom	$\chi 2$ statistic for Hardy-Weinberg Probability test	Probability	Fwc(st)
rs659366	8	14.5	.07 (ns)	.0108

Table 3.3 $\chi 2$-distributed Test Statistics for Independence for UCP2 –866 rs659366.

SNP	Sample size	Trend test for additive allele using only Obese and Lean	Non-additive χ^2 statistic (4df)	G-test with William's correction (4df)
rs659366	246	9.0*10^-3 (ns)	1.86 (ns)	1.78 (ns)

HLA haplotype and diabetes in the Pima Indian population and in Blum *et al.*, study on the dopamine DRD2 allele and alcoholism (Blum *et al.*, 1990; Gelernter *et al.*, 1993; Knowler *et al.*, 1988). These studies may be the exception not the rule. Wacholder *et al.*, showed that population stratification bias is rare and could be accounted for by proper inquires to people's ethnic origins (Wacholder *et al.*, 2000). We must genotype greater than 500 Alaska Native's with more refined phenotypes, correcting for variables affecting obesity, in order to support or refute the UCP2 –866 A allele association with decreased risk of obesity in the Alaska Native population.

It is interesting that the Alaska Native sample set had "G" as the minor allele, while "A" is the minor allele in the multi-ethnic sample set. A study done by Sesti *et al.*, suggested that the –866 polymorphism plays a role in the variation of insulin secretion; they found that islets from AA homozygotes had a lower insulin response to glucose administration compared to the other genotypes (Sesti *et al.*, 2003). Since "A" is the major allele in the Alaska Native sample set, it might prove illuminating to test whether the same association exists.

CHAPTER 4: TESTING LEPR A/G RS1137101 FOR ASSOCIATION WITH BMI IN WESTERN ALASKA NATIVES

INTRODUCTION

LEPR has been implemented as a candidate gene for obesity by both transmission-linkage and case-control studies (Chagnon *et al.*, 1999; Chagnon *et al.*, 2000; Mattevi *et al.*, 2002). A mutation in LEPR has been shown to cause severe obesity in humans and several *Lepr* mutation mouse models exhibit early-onset obesity (Clement *et al.*, 1998). LEPR has been cloned and mapped to 1p31 (Chung *et al.*, 1996; Tartaglia *et al.*, 1995).

The common polymorphism, LEPR Gln(CAG)223Arg(CGG), rs1137101, has exhibited both linkage and case-control association with BMI in multiple human studies (Chagnon *et al.*, 1999; Chagnon *et al.*, 2000; Mattevi *et al.*, 2002). The polymorphism lies in a Cytokine domain that is a possible binding site for leptin (Tartaglia *et al.*, 1995). Chagnon *et al.* hypothesized that altering the amino acid at position 223 could change the leptin receptor's signaling capacity (Chagnon *et al.*, 2000). In this case-control study rs1137101, Gln223Arg, was tested for its association with BMI in the Western Alaska Native population. The null hypothesis is that the genotype is independent of BMI category.

METHODS

75 Alaska Native samples were PCR amplified and sequenced with LEPR exon 4 reverse primer, described in Chapter 1. Hardy-Weinberg was tested with Genepop. A trend test (additive), chi-square (non additive), and G test of independence (non additive) were used to test the null hypothesis that the genotype at the locus is independent of BMI group.

RESULTS

Hardy-Weinberg could not be rejected with a probability of .69 that the difference between the observed and expected was due to chance (Table 4.1). The null hypothesis that the locus is independent of BMI could not be rejected in any of the chi-square test statistics (Table 4.1). The minor allele, A, in the sample set had a frequency of .18, considerably lower than other reported minor allele frequencies at the locus (Table 4.2).

DISCUSSION

No departures from the Hardy-Weinberg assumptions were large enough to be detected by the chi-square test. This lends support that, with respect to this locus, evolutionary forces are negligible and that the population is randomly mating. The minor allele frequency of the sample set is considerably lower than those reported by Celera and dbSNP for Caucasian, African American, and multi-ethnic groups; this highlights the uniqueness of the Alaska Native population and reiterates the importance of continuing genetic study. The SNP is still considered to be "common" even though the Alaska Native samples had a lower minor allele frequency than the other groups. The definition of a common allele is somewhat arbitrary and ranges from a minor allele frequency >1% to a minor allele frequency >10% (Cargill *et al.*, 1999; Risch, 2000).

The tests for association could not reject the null hypothesis that BMI category is independent from the locus. This may mean that the locus is not associated with BMI or that there was not enough power to detect an association in the sample set. Since the sample set was de-identified, variables known to affect BMI could not be corrected. The sample size was 72 which is much less than Long and Langley's recommended 500 or greater (Long and Langley, 1999). Future studies will greatly benefit from increased sample size, sample information pertinent to BMI, and acquisition of refined phenotypes. LEPR rs1137101 remains a candidate locus for contributing to the development of obesity. Continuing

Table 4.1 LEPR G/A, rs1137101, χ2-distributed Test Statistics for Independence and Hardy-Weinberg equilibrium.

SNP	Sample size	Hardy-Weinberg Probability (1df)	Trend test for additive allele using only Obese and Lean	Non-additive χ^2 statistic (4df)	G-test with William's correction (4df)
rs1137101	72	.69 (ns)	1.01*10^-6 (ns)	7.32 (ns)	7.57 (ns)

Table 4.2 LEPR G/A SNP, rs1137101, Minor Allele Frequencies.

	Caucasian (Celera)	African American (Celera)	Multi-ethnic (dbSNP)	Alaska Native
Minor allele (A) frequency	.48	.42	.34	.18

studies in the Alaska Native population, as well as others, will help determine how much of a role the locus has in the etiology of obesity.

GENERAL CONCLUSION

SNP DISCOVERY

Fewer SNPs were "discovered" than expected based on previous studies of SNPs in coding regions. This is most likely attributed to the stochastic nature of sampling a small number of base pairs (5043), sampling a single population having increased homogeneity, and missing SNPs. Increasing the number of coding base pairs screened would likely cause the ratio of SNP per coding base pair screened to approach those reported by Cargill *et al.* and Stephens *et al.* (Cargill *et al.*, 1999; Stephens *et al.*, 2001).

SNPs found were previously characterized in other populations. This is not surprising given that most human variation is cosmopolitan (ancient origin) and only 20% is either population specific or shared between related populations. Therefore, it would be more economical to use common known SNPs for fine mapping over candidate gene regions or for functional mapping. Since we've shown that Alaska Natives can have very different allele frequencies than other populations, the following question arises: How do we know SNPs that are common in other populations are common in the Alaska Native population? All 5 SNPs genotyped in the Alaska Native population had minor allele frequencies >5%. As was mentioned in chapter 2, the definition of "common" is somewhat arbitrary, ranging from minor allele frequencies >1% to >10%. Researchers often use the intermediate definition of "common" as meaning the minor allele frequency is >5%. If we adopt this same definition, then all 5 SNPs studied are common in the Alaska Native population, despite being different in frequency than other populations.

HARDY-WEINBERG, FST, AND TESTS FOR ASSOCIATION WITH BMI

Hardy-Weinberg could not be rejected for all loci studied, and rs5573, rs5574, and UCP2 -866 indicated little genetic differentiation between subpopulations by Fwc(st) values <.05 (Wright, 1978). The null hypothesis that genotype is independent of BMI group could not be rejected for any of the loci. In particular, previously reported associations with BMI for

the UCP2 promoter SNP and the LEPR SNP were not supported using this sample set. This could mean that there is no association between the SNPs and BMI in the Alaska Native population or there was not enough power to detect an association. To increase power, variables known to affect BMI, such as age and sex, would need to be corrected for and the sample size would need to be increased (Long and Langley, 1999).

NON-REPLICATION OF CASE CONTROL ASSOCIATION STUDIES

Non-replication of case-control association studies is not surprising. Hirschhorn *et al.* reviewed 600 association studies that reported positive associations and found that only 6 had been consistently replicated three or more times (Hirschhorn *et al.*, 2002). They then discussed possible causes of false positive and false negative association reports. Possible causes of false positive results were listed as population stratification from ethnic admixture, variation of LD with causative variant, and gene-gene or gene-environment interactions specific to a certain population. Hirschhorn *et al.* said that underpowered studies along with weak genetic effects are most likely the cause for falsely negative associations. The results of our study could very well lie in the category of falsely negative association studies for those reasons. Due to high non-replication levels of case-case control studies, *Nature* adopted stringent publication guidelines: large sample sizes, associations must make biological sense, small P values, initial study must be replicated, and case-control results must be supported by family based studies.

Few research groups are likely to fulfill all of *Nature's* association study guidelines. This does not mean that case-control association studies should not be done. Rather, it emphasizes the need for collaborative research. It also suggests that a database containing all association results, positive and negative, and enough information to perform meta-analyses would greatly increase the value of case-case control studies, regardless of the size of the study or number of guidelines fulfilled.

UNIQUE ALLELE FREQUENCIES

The NPY exon 2 SNP was the only SNP having a minor allele frequency close to those reported in another sample set. The greatest difference in minor alleles was observed in the UCP2 –866 SNP, where the Alaska Native sample set had an alternate minor allele than the dbSNP multi-ethnic panel. The unique allele frequencies found in this Alaska Native sample set confirm the need for further genetic studies. The most probable explanation for the difference in allele frequencies is that the Alaska Natives underwent a series of population bottlenecks created by founder effects and/or by numerous epidemics. Population bottlenecks have been shown to increase the incidence of rare diseases in several human populations (Hartl and Clark, 1997). The vastly different allele frequencies in the Alaska Native population could contribute to the susceptibility of certain diseases. If a study could control for the environment, allele frequencies could be compared between populations to test whether the unique allele frequencies found in Alaska Natives are able to explain increased risk of obesity.

METHODS OF SNP GENOTYPING

Different methods of SNP genotyping worked better or worse depending on the locus. Applied Biosystem's assays on demand 5' nuclease reactions were the most convenient and easy to use because no optimization is required. SBEX and DHPLC require optimization, and sequencing is too expensive for high throughput SNP genotyping. The take home message is that it is beneficial to have more than one SNP genotyping method. The International HapMap project illustrates this point by using five different genotyping technologies; "a SNP assay that fails on one platform may be developed successfully using another method in order to fill a gap in the HapMap" (Gibbs *et al.*, 2003).

INDIRECT VS. DIRECT CASE-CONTROL ASSOCIATION STUDIES

Case-control studies looking for loci associated with the variation of a complex trait may be carried out in a direct or indirect manner (Collins *et al.*, 1997). Direct case-control

association studies test only functional variants for disease association. Functional variants are those that change the protein or the expression of the gene. Indirect studies use dense regularly spaced SNPs to cover a region of interest. A direct approach using common SNPs to screen candidate genes would be ideal. Advantages of choosing functional SNPs are that they are more likely to be responsible for the phenotypic variation of the candidate gene, they could narrow the number of SNPs needed to be genotyped, and that they are not wholly dependent on LD. Indirect studies are fully dependent upon LD; LD is highly variable within the human genome and between populations, creating debate over the number of SNPs required (Kruglyak, 1999; Risch, 2000). Additionally, LD is usually not complete so the genetic effect is diluted, making it more difficult to detect loci affecting the variation of a trait (Cardon and Bell, 2001). Ohashi and Tokunaga tested the power of indirect vs. direct studies to find disease loci, and they found that indirect studies have severely reduced statistical power (Ohashi and Tokunaga, 2001).

In future studies I would choose a direct case-control association approach using common functional SNPs. Functional SNPs can be divided into two categories: those causing amino acid changes and those putatively altering gene expression. SNPs potentially altering gene expression would be higher priority than those altering amino acids. Several obesity and diabetes studies have reported associations between common promoter region SNPs and variation in phenotype (Hoffstedt *et al.*, 2002; Muller *et al.*, 2003; Wang *et al.*, 2002; Wang *et al.*, 2004). SNPs causing amino acid changes are of lower priority because conservative amino acid changes are unlikely to alter the function of the protein, while large amino acid changes are likely to be rare and deleterious. SNPs that may alter expression would be further prioritized according to whether they are located in conserved promoter regions and how many putative transcription factor binding sites they alter. A direct case-control method using common SNPs potentially altering candidate gene expression has three main benefits over an indirect method: it reduces the number of SNPs assayed per gene, it is not wholly dependent on LD, and positive associations are testable for biological significance using functional assays.

Before indirect or direct case-control association methods are applied it would be useful to conduct a genome scan in the Alaska Native population to determine the broad regions of the genome that are linked to obesity in the Alaska Native population. Previously characterized candidate obesity genes lying close to linkage peaks would be primary targets of study, as well as genes making biological sense.

SUMMATION

In summation, a completely anonymized sample set study design is ill advised. Too much sample information is lost leading to decreased statistical power. Case-control association studies of complex diseases often already suffer from a lack power due to there being many genes of weak effect and from inadequate sample sizes. If for ethical reasons a sample set must be anonymized the most refined phenotype should be kept, as long as it does not compromise the anonymity of the samples. A more refined phenotype would allow some statistical corrections for variables affecting the phenotype of interest

LITERATURE CITED

Adeyemo, A., Luke, A., Cooper, R., Wu, X., Tayo, B., Zhu, X., Rotimi, C., Bouzekri, N. & Ward, R. A Genome-Wide Scan for Body Mass Index among Nigerian Families. *Obes Res* **11**, 266-273 (2003).

Adler, A. I., Boyko, E. J., Schraer, C. D. & Murphy, N. J. The negative association between traditional physical activities and the prevalence of glucose intolerance in Alaska Natives. *Diabet Med* **13**, 555-60 (1996).

Allison, D. B., Kaprio, J., Korkeila, M., Koskenvuo, M., Neale, M. C. & Hayakawa, K. The heritability of body mass index among an international sample of monozygotic twins reared apart. *Int J Obes Relat Metab Disord* **20**, 501-6 (1996).

Arya, R., Duggirala, R., Jenkinson, C. P., Almasy, L., Blangero, J., O'Connell, P. & Stern, M. P. Evidence of a novel quantitative-trait locus for obesity on chromosome 4p in mexican americans. *Am J Hum Genet* **74**, 272-82 (2004).

Barsh, G. S., Farooqi, I. S. & O'Rahilly, S. Genetics of body-weight regulation. *Nature* **404**, 644-51 (2000).

Billington, C. J., Briggs, J. E., Grace, M. & Levine, A. S. Effects of intracerebroventricular injection of neuropeptide Y on energy metabolism. *Am J Physiol* **260**, R321-7 (1991).

Bjorbaek, C., Elmquist, J. K., Frantz, J. D., Shoelson, S. E. & Flier, J. S. Identification of SOCS-3 as a potential mediator of central leptin resistance. *Mol Cell* **1**, 619-25 (1998).

Blum, K., Noble, E. P., Sheridan, P. J., Montgomery, A., Ritchie, T., Jagadeeswaran, P., Nogami, H., Briggs, A. H. & Cohn, J. B. Allelic association of human dopamine D2 receptor gene in alcoholism. *Jama* **263**, 2055-60 (1990).

Boss, O., Hagen, T. & Lowell, B. B. Uncoupling proteins 2 and 3: potential regulators of mitochondrial energy metabolism. *Diabetes* **49**, 143-56 (2000).

Bouchard, C., Perusse, L., Chagnon, Y. C., Warden, C. & Ricquier, D. Linkage between markers in the vicinity of the uncoupling protein 2 gene and resting metabolic rate in humans. *Hum Mol Genet* **6**, 1887-9 (1997).

Bray, M. S., Boerwinkle, E. & Hanis, C. L. Linkage analysis of candidate obesity genes among the Mexican-American population of Starr County, Texas. *Genet Epidemiol* **16**, 397-411 (1999).

Bray, M. S., Boerwinkle, E. & Hanis, C. L. Sequence variation within the neuropeptide Y gene and obesity in Mexican Americans. *Obes Res* **8**, 219-26 (2000).

Budowle, B., Chidambaram, A., Strickland, L., Beheim, C. W., Taft, G. M. & Chakraborty, R. Population studies on three Native Alaska population groups using STR loci. *Forensic Sci Int* **129**, 51-7 (2002).

Butler, M. G., Hedges, L., Hovis, C. L. & Feurer, I. D. Genetic variants of the human obesity (OB) gene in subjects with and without Prader-Willi syndrome: comparison with body mass index and weight. *Clin Genet* **54**, 385-93 (1998).

Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nat Rev Genet* **2**, 91-9 (2001).

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. Q. & Lander, E. S. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* **22**, 231-8 (1999).

Chagnon, Y. C., Chen, W. J., Perusse, L., Chagnon, M., Nadeau, A., Wilkison, W. O. & Bouchard, C. Linkage and association studies between the melanocortin receptors 4 and 5 genes and obesity-related phenotypes in the Quebec Family Study. *Mol Med* **3**, 663-73 (1997).

Chagnon, Y. C., Chung, W. K., Perusse, L., Chagnon, M., Leibel, R. L. & Bouchard, C. Linkages and associations between the leptin receptor (LEPR) gene and human body composition in the Quebec Family Study. *Int J Obes Relat Metab Disord* **23**, 278-86 (1999).

Chagnon, Y. C., Wilmore, J. H., Borecki, I. B., Gagnon, J., Perusse, L., Chagnon, M., Collier, G. R., Leon, A. S., Skinner, J. S., Rao, D. C. & Bouchard, C. Associations between the leptin receptor gene and adiposity in middle-aged Caucasian males from the HERITAGE family study. *J Clin Endocrinol Metab* **85**, 29-34 (2000).

Chen, X., Levine, L. & Kwok, P. Y. Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res* **9**, 492-8 (1999).

Chua, S. C., Jr., Brown, A. W., Kim, J., Hennessey, K. L., Leibel, R. L. & Hirsch, J. Food deprivation and hypothalamic neuropeptide gene expression: effects of strain background and the diabetes mutation. *Brain Res Mol Brain Res* **11**, 291-9 (1991).

Chung, W. K., Power-Kehoe, L., Chua, M. & Leibel, R. L. Mapping of the OB receptor to 1p in a region of nonconserved gene order from mouse and rat to human. *Genome Res* **6**, 431-8 (1996).

Clement, K., Vaisse, C., Lahlou, N., Cabrol, S., Pelloux, V., Cassuto, D., Gourmelen, M., Dina, C., Chambaz, J., Lacorte, J. M., Basdevant, A., Bougneres, P., Lebouc, Y., Froguel, P. & Guy-Grand, B. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398-401 (1998).

Collins, F. S., Guyer, M. S. & Chakravarti, A. Variations on a Theme: Cataloging Human DNA Sequence Variation. *Science* **278**, 1580-1581 (1997).

Comuzzie, A. G., Hixson, J. E., Almasy, L., Mitchell, B. D., Mahaney, M. C., Dyer, T. D., Stern, M. P., MacCluer, J. W. & Blangero, J. A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat Genet* **15**, 273-6 (1997).

Cone, R. D., Lu, D., Koppula, S., Vage, D. I., Klungland, H., Boston, B., Chen, W., Orth, D. N., Pouton, C. & Kesterson, R. A. The melanocortin receptors: agonists, antagonists, and the hormonal control of pigmentation. *Recent Prog Horm Res* **51**, 287-317; discussion 318 (1996).

Dolinski, K., Balakrishnan, R., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hong, E. L., Issel-Tarver, L., Sethuraman, A., Theesfeld, C. L., Binkley, G., Lane, C., Schroeder, M., Dong, S., Weng, S., Andrada, R., Botstein, D. & Cherry, J. M. (2003).

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. "Touchdown" PCR to eliminate spurious priming during gene amplification. *Nucleic Acids Res.* **19**, 4008 (1991).

Echtay, K. S., Murphy, M. P., Smith, R. A., Talbot, D. A. & Brand, M. D. Superoxide activates mitochondrial uncoupling protein 2 from the matrix side. Studies using targeted antioxidants. *J Biol Chem* **277**, 47129-35 (2002).

Erickson, J. C., Hollopeter, G. & Palmiter, R. D. Attenuation of the obesity syndrome of ob/ob mice by the loss of neuropeptide Y. *Science* **274**, 1704-7 (1996).

Esterbauer, H., Schneitler, C., Oberkofler, H., Ebenbichler, C., Paulweber, B., Sandhofer, F., Ladurner, G., Hell, E., Strosberg, A. D., Patsch, J. R., Krempler, F. & Patsch, W. A common polymorphism in the promoter of UCP2 is associated with decreased risk of obesity in middle-aged humans. *Nat Genet* **28**, 178-83 (2001).

Evans, D., Minouchehr, S., Hagemann, G., Mann, W. A., Wendt, D., Wolf, A. & Beisiegel, U. Frequency of and interaction between polymorphisms in the beta3-adrenergic receptor and in uncoupling proteins 1 and 2 and obesity in Germans. *Int J Obes Relat Metab Disord* **24**, 1239-45 (2000).

Evans, D., Wolf, A. M., Nellessen, U., Ahle, S., Kortner, B., Kuhlmann, H. W. & Beisiegel, U. Association between polymorphisms in candidate genes and morbid obesity. *Int J Obes Relat Metab Disord* **25 Suppl 1**, S19-21 (2001).

French, S. A., Story, M. & Jeffery, R. W. Environmental influences on eating and physical activity. *Annu Rev Public Health* **22**, 309-35 (2001).

Gelernter, J., Goldman, D. & Risch, N. The A1 allele at the D2 dopamine receptor gene and alcoholism. A reappraisal. *Jama* **269**, 1673-7 (1993).

Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., Ch'ang, L. Y., Huang, W., Liu, B., Shen, Y., Tam, P. K., Tsui, L. C., Waye, M. M., Wong, J. T., Zeng, C., Zhang, Q., Chee, M. S., Galver, L. M., Kruglyak, S., Murray, S. S., Oliphant, A. R., Montpetit, A., Hudson, T. J., Chagnon, F., Ferretti, V., Leboeuf, M., Phillips, M. S., Verner, A., Kwok, P. Y., Duan, S., Lind, D. L., Miller, R. D., Rice, J. P., Saccone, N. L., Taillon-Miller, P., Xiao, M., Nakamura, Y., Sekine, A., Sorimachi, K., Tanaka, T., Tanaka, Y., Tsunoda, T., Yoshino, E., Bentley, D. R., Deloukas, P., Hunt, S., Powell, D., Altshuler, D., Gabriel, S. B., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Aniagwu, T., Marshall, P. A., Matthew, O., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Stein, L. D., Cunningham, F., Kanani, A., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Donnelly, P., Marchini, J., McVean, G. A., Myers, S. R., Cardon, L. R., Abecasis, G. R., Morris, A., Weir, B. S., Mullikin, J. C., Sherry, S. T., Feolo, M., Daly, M. J., Schaffner, S. F., Qiu, R., Kent, A., Dunston, G. M., Kato, K., Niikawa, N., Knoppers, B. M., Foster, M. W., Clayton, E. W., Wang, V. O., Watkin, J., Sodergren, E., Weinstock, G. M., Wilson, R. K., Fulton, L. L., Rogers, J., Birren, B. W., Han, H., Wang, H., Godbout, M., Wallenburg, J. C., L'Archeveque, P., Bellemare, G., Todani, K., Fujita, T., Tanaka, S., Holden, A. L., Lai, E. H., Collins, F. S., Brooks, L. D., McEwen, J. E., Guyer, M. S., Jordan, E., Peterson, J. L., Spiegel, J., Sung, L. M., Zacharia, L. F., Kennedy, K., Dunn, M. G., Seabrook, R., Shillito, M., Skene, B., Stewart, J. G., Valle, D. L., Jorde, L. B., Cho, M. K., Duster, T., Jasperse, M., Licinio, J., Long, J. C., Ossorio, P. N., Spallone, P., Terry, S. F., Lander, E. S., Nickerson, D. A., Boehnke, M., Douglas, J. A., Hudson, R. R., Kruglyak, L. & Nussbaum, R. L. The International HapMap Project. *Nature* **426**, 789-96 (2003).

Gimeno, R. E., Dembski, M., Weng, X., Deng, N., Shyjan, A. W., Gimeno, C. J., Iris, F., Ellis, S. J., Woolf, E. A. & Tartaglia, L. A. Cloning and characterization of an uncoupling protein homolog: a potential molecular mediator of human thermogenesis. *Diabetes* **46**, 900-6 (1997).

Halushka, M. K., Fan, J. B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. & Chakravarti, A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* **22**, 239-47 (1999).

Hartl, D. & Clark, A. *Principles of Population Genetics* (Sinauer Associates, Inc., Sunderland, 1997).

Heller, C. & Scott, E. M. (US Dept. of Health, Education and Welfare, Washington, DC, 1967).

Hill, J. O. & Peters, J. C. Environmental Contributions to the Obesity Epidemic. *Science* **280**, 1371-1374 (1998).

Hinney, A., Hohmann, S., Geller, F., Vogel, C., Hess, C., Wermter, A. K., Brokamp, B., Goldschmidt, H., Siegfried, W., Remschmidt, H., Schafer, H., Gudermann, T. & Hebebrand, J. Melanocortin-4 receptor gene: case-control study and transmission disequilibrium test confirm that functionally relevant mutations are compatible with a major gene effect for extreme obesity. *J Clin Endocrinol Metab* **88**, 4258-67 (2003).

Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet Med* **4**, 45-61 (2002).

Hoffstedt, J., Andersson, I. L., Persson, L., Isaksson, B. & Arner, P. The common -675 4G/5G polymorphism in the plasminogen activator inhibitor -1 gene is strongly associated with obesity. *Diabetologia* **45**, 584-7 (2002).

Huszar, D., Lynch, C. A., Fairchild-Huntress, V., Dunmore, J. H., Fang, Q., Berkemeier, L. R., Gu, W., Kesterson, R. A., Boston, B. A., Cone, R. D., Smith, F. J., Campfield, L. A., Burn, P. & Lee, F. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131-41 (1997).

Jaburek, M., Varecha, M., Gimeno, R. E., Dembski, M., Jezek, P., Zhang, M., Burn, P., Tartaglia, L. A. & Garlid, K. D. Transport function and regulation of mitochondrial uncoupling proteins 2 and 3. *J Biol Chem* **274**, 26003-7 (1999).

Jiang, R., Duan, J., Windemuth, A., Stephens, J. C., Judson, R. & Xu, C. Genome-wide evaluation of the public SNP databases. *Pharmacogenomics* **4**, 779-89 (2003).

Karvonen, M. K., Koulu, M., Pesonen, U., Uusitupa, M. I., Tammi, A., Viikari, J., Simell, O. & Ronnemaa, T. Leucine 7 to proline 7 polymorphism in the preproneuropeptide Y is associated with birth weight and serum triglyceride concentration in preschool aged children. *J Clin Endocrinol Metab* **85**, 1455-60 (2000).

Knapp, B. & Panruk, P. (Yukon Kuskokwim Health Corporation Publishing, Bethel, Alaska, 1978).

Knowler, W. C., Williams, R. C., Pettitt, D. J. & Steinberg, A. G. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* **43**, 520-6 (1988).

Kotz, C. M., Briggs, J. E., Grace, M. K., Levine, A. S. & Billington, C. J. Divergence of the feeding and thermogenic pathways influenced by NPY in the hypothalamic PVN of the rat. *Am J Physiol Regul Integr Comp Physiol* **275**, R471-477 (1998).

Krempler, F., Esterbauer, H., Weitgasser, R., Ebenbichler, C., Patsch, J. R., Miller, K., Xie, M., Linnemayr, V., Oberkofler, H. & Patsch, W. A functional polymorphism in the

promoter of UCP2 enhances obesity risk but reduces type 2 diabetes risk in obese middle-aged humans. *Diabetes* **51**, 3331-5 (2002).

Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**, 139-44 (1999).

Le Fur, S., Le Stunff, C., Dos Santos, C. & Bougneres, P. The common -866 g/a polymorphism in the promoter of uncoupling protein 2 is associated with increased carbohydrate and decreased lipid oxidation in juvenile obesity. *Diabetes* **53**, 235-9 (2004).

Lee, J. H., Reed, D. R. & Price, R. A. Familial risk ratios for extreme obesity: implications for mapping human obesity genes. *Int J Obes Relat Metab Disord* **21**, 935-40 (1997).

Li, W. D., Reed, D. R., Lee, J. H., Xu, W., Kilker, R. L., Sodam, B. R. & Price, R. A. Sequence variants in the 5' flanking region of the leptin gene are associated with obesity in women. *Ann Hum Genet* **63 (Pt 3)**, 227-34 (1999).

Long, A. D. & Langley, C. H. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res* **9**, 720-31 (1999).

Loos, R. J. & Bouchard, C. Obesity--is it a genetic disorder? *J Intern Med* **254**, 401-25 (2003).

Luke, A., Guo, X., Adeyemo, A. A., Wilks, R., Forrester, T., Lowe, W., Jr., Comuzzie, A. G., Martin, L. J., Zhu, X., Rotimi, C. N. & Cooper, R. S. Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. *Int J Obes Relat Metab Disord* **25**, 1034-41 (2001).

Maes, H. H., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* **27**, 325-51 (1997).

Mattevi, V. S., Zembrzuski, V. M. & Hutz, M. H. Association analysis of genes involved in the leptin-signaling pathway with obesity in Brazil. *Int J Obes Relat Metab Disord* **26,** 1179-85 (2002).

Michael, H. *Lietenant Zagoskin's Travels in Russian America* (ed. HN, M.) (University of Toronto Press, 1967).

Millipore. Tech Note TN053 (2004).

Mouratoff, G. J., Carroll, N. V. & Scott, E. M. Diabetes mellitus in Eskimos. *Jama* **199,** 107-12 (1967).

Mouratoff, G. J. & Scott, E. M. Diabetes mellitus in Eskimos after a decade. *Jama* **226,** 1345-6 (1973).

Muller, Y. L., Bogardus, C., Beamer, B. A., Shuldiner, A. R. & Baier, L. J. A functional variant in the peroxisome proliferator-activated receptor gamma2 promoter is associated with predictors of obesity and type 2 diabetes in Pima Indians. *Diabetes* **52,** 1864-71 (2003).

Murphy, N. J., Schraer, C. D., Thiele, M. C., Boyko, E. J., Bulkow, L. R., Doty, B. J. & Lanier, A. P. Dietary change and obesity associated with glucose intolerance in Alaska Natives. *J Am Diet Assoc* **95,** 676-82 (1995).

O'Donohue, T. L., Chronwall, B. M., Pruss, R. M., Mezey, E., Kiss, J. Z., Eiden, L. E., Massari, V. J., Tessel, R. E., Pickel, V. M., DiMaggio, D. A. & et al. Neuropeptide Y and peptide YY neuronal and endocrine systems. *Peptides* **6,** 755-68 (1985).

Ohashi, J. & Tokunaga, K. The power of genome-wide association studies of complex disease genes: statistical limitations of indirect approaches using SNP markers. *J Hum Genet* **46,** 478-82 (2001).

Palmer, L. J. & Cookson, W. O. Using single nucleotide polymorphisms as a means to understanding the pathophysiology of asthma. *Respir Res* **2**, 102-12 (2001).

Perrin, F. Polarization de la lumiere de fluorescence. Vie moyenne de molecules dans l'etat excite. *J. Phys. Radium* **7**, 390-401 (1926).

Qiagen. *Qiaquick Multiwell PCR Purification Handbook* (Qiagen, Valencia, 2002).

Qiagen. *QIAquick Spin Handbook* (Qiagen, Valcencia, 2002).

Risch, N. J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847-56 (2000).

Schmitt, T. J., Robinson, M. L. & Doyle, J. Single Nucleotide Polymorphism (SNP), Insertion & Deletion Detection on the WAVE Nucleic Acid Fragment Analysis System. *Transgenomic Application Note* **112**, 1-4.

Schwartz, M. W., Woods, S. C., Porte, D., Jr., Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404**, 661-71 (2000).

Scott, E. M. & Griffith, I. V. Diabetes mellitus in Eskimos. *Metabolism* **6**, 320-5 (1957).

Scott, E. M. & Wright, R. C. Genetic diversity of Central Yupik Eskimos. *Hum Biol* **55**, 409-15 (1983).

Sesti, G., Cardellini, M., Marini, M. A., Frontoni, S., D'Adamo, M., Del Guerra, S., Lauro, D., De Nicolais, P., Sbraccia, P., Del Prato, S., Gambardella, S., Federici, M., Marchetti, P. &

Lauro, R. A common polymorphism in the promoter of UCP2 contributes to the variation in insulin secretion in glucose-tolerant subjects. *Diabetes* **52**, 1280-3 (2003).

Shields, G. F., Schmiechen, A. M., Frazier, B. L., Redd, A., Voevoda, M. I., Reed, J. K. & Ward, R. H. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am J Hum Genet* **53**, 549-62 (1993).

Sokal, F. & Rohlf, J. *Biometry* (W.H. Freeman and Company, New York, 1995).

Stanley, B. G., Kyrkouli, S. E., Lampert, S. & Leibowitz, S. F. Neuropeptide Y chronically injected into the hypothalamus: a powerful neurochemical inducer of hyperphagia and obesity. *Peptides* **7**, 1189-92 (1986).

Stanley, B. G. & Leibowitz, S. F. Neuropeptide Y: stimulation of feeding and drinking by injection into the paraventricular nucleus. *Life Sci* **35**, 2635-42 (1984).

Stephens, J. C., Schneider, J. A., Tanguay, D. A., Choi, J., Acharya, T., Stanley, S. E., Jiang, R., Messer, C. J., Chew, A., Han, J. H., Duan, J., Carr, J. L., Lee, M. S., Koshy, B., Kumar, A. M., Zhang, G., Newell, W. R., Windemuth, A., Xu, C., Kalbfleisch, T. S., Shaner, S. L., Arnold, K., Schulz, V., Drysdale, C. M., Nandabalan, K., Judson, R. S., Ruano, G. & Vovis, G. F. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**, 489-93 (2001).

Stephens, T. W., Basinski, M., Bristow, P. K., Bue-Valleskey, J. M., Burgett, S. G., Craft, L., Hale, J., Hoffmann, J., Hsiung, H. M., Kriauciunas, A. & et al. The role of neuropeptide Y in the antiobesity action of the obese gene product. *Nature* **377**, 530-2 (1995).

Stunkard, A. J., Harris, J. R., Pedersen, N. L. & McClearn, G. E. The body-mass index of twins who have been reared apart. *N Engl J Med* **322**, 1483-7 (1990).

Tartaglia, L. A., Dembski, M., Weng, X., Deng, N., Culpepper, J., Devos, R., Richards, G. J., Campfield, L. A., Clark, F. T., Deeds, J. & et al. Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263-71 (1995).

Wacholder, S., Rothman, N. & Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* **92**, 1151-8 (2000).

Wang, H., Chu, W. S., Hemphill, C. & Elbein, S. C. Human resistin gene: molecular scanning and evaluation of association with insulin sensitivity and type 2 diabetes in Caucasians. *J Clin Endocrinol Metab* **87**, 2520-4 (2002).

Wang, H., Chu, W. S., Lu, T., Hasstedt, S. J., Kern, P. A. & Elbein, S. C. Uncoupling protein-2 polymorphisms in type 2 diabetes, obesity, and insulin secretion. *Am J Physiol Endocrinol Metab* **286**, E1-7 (2004).

Williams, D. A. Improved likelihood ratio tests for complete contingency tables. *Biometrika* **63**, 33-37 (1976).

Wright, S. *Evolution and the Genetics of Populations.* (University of Chicago Press, Chicago, 1978).

Zarjevski, N., Cusin, I., Vettor, R., Rohner-Jeanrenaud, F. & Jeanrenaud, B. Chronic intracerebroventricular neuropeptide-Y administration to normal rats mimics hormonal and metabolic changes of obesity. *Endocrinology* **133**, 1753-8 (1993).

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L. & Friedman, J. M. Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425-32 (1994).